

On the Approximability of Orthogonal Order Preserving Layout Adjustment

Sayan Bandyapadhyay, Santanu Bhowmick, and Kasturi Varadarajan

Department of Computer Science
University of Iowa, Iowa City, USA

Abstract. Given an initial placement of a set of rectangles in the plane, we consider the problem of finding a disjoint placement of the rectangles that minimizes the area of the bounding box and preserves the orthogonal order i.e. maintains the sorted ordering of the rectangle centers along both x -axis and y -axis with respect to the initial placement. This problem is known as *Layout Adjustment for Disjoint Rectangles* (LADR). It was known that LADR is NP-hard, but only heuristics were known for it. We show that a certain decision version of LADR is \mathbb{APX} -hard, and give a constant factor approximation for LADR.

1 Introduction

Graphs are often used to visualize relationships between entities in diverse fields such as software engineering (e.g. UML diagrams), VLSI (circuit schematics) and biology (e.g. biochemical pathways) [13]. For many such applications, treating graph nodes as points is insufficient, since each node may have a corresponding label explaining its significance. The presence of labels may lead to node overlapping. For the typical user, an uncluttered layout is more important than the amount of information presented [21]. For complex graphs, it is tedious to create meaningful layouts by hand, which has led to algorithms for layout generation.

Layout generation algorithms typically take a combinatorial description of a graph, and return a corresponding layout. Nodes are usually represented by boxes, and edges by lines connecting the boxes. For simplicity, the edges of the graph are ignored while creating the modified layout. In some interactive systems, modifications to the graph may happen in multiple stages. The layout must be adjusted after each alteration (if new nodes added overlap existing nodes), such that the display area is minimized. If we use layout creation algorithms after each iteration, we may get a layout that is completely different from the previous layout, which may destroy the ‘mental map’ of the user who is interacting with the system. Thus, we need an additional constraint in the form of maintaining some property of the layout, which would be equivalent to preserving the mental map. Eades et al. [6] defined *orthogonal ordering* as one of the key properties that should be maintained in an adjusted layout to preserve the user’s mental map. Two layouts of a graph have the same orthogonal ordering if the horizontal and vertical ordering of the nodes are identical in both layouts.

We now state the problem studied in this paper, which involves laying out rectangles that represent the nodes in the graph being adjusted. We are given

a set of rectangles R (each $r_i \in R$ is defined by an ordered pair, $r_i = (w_i, h_i)$, denoting its width and height respectively) and an initial layout λ^{in} . A layout consists of an assignment $\lambda : R \rightarrow \mathbb{R}^2$ of coordinates to the centers of rectangles in R . The goal is to find a layout in which no two rectangles intersect and orthogonal ordering of the rectangle centers w.r.t λ^{in} is maintained, while minimizing the area of the bounding box of the layout. We refer to this problem as *Layout Adjustment for Disjoint Rectangles* (LADR). Note that R is really a set of rectangle dimensions, and not a set of rectangles. Nevertheless, we will refer to R as a set of rectangles. See [Section 2](#) for a more leisurely problem statement.

1.1 Previous Work

The concept of a mental map was introduced in [6], along with three quantitative models representing it - orthogonal ordering, proximity relations and topology. A framework for analyzing the various models of a mental map was presented in [4], which determined that orthogonal ordering constraint was the best metric for comparing different drawings of the same graph. A user study designed to evaluate human perceptions of similarity amongst two sets of drawings was given in [5], in which orthogonal ordering constraints received the highest rankings.

There has been a lot of work done using the concept of preserving mental maps. LADR was first introduced in [18], in which the authors described the *Force-Scan* (FS) algorithm. FS scans for overlapping nodes in both horizontal and vertical directions, and separates two intersecting nodes by “forcing” them apart along the line connecting the centers of the two nodes, while ensuring that the nodes being forced apart do not intersect any additional nodes in the layout. In [12], a modification of FS was presented (FS’), which resulted in a more compact layout than FS. Another version of FS algorithm, called the *Force-Transfer* (FT) algorithm, was given in [14]. For any two overlapping nodes, denote the vertical distance to be moved to remove the overlap as d_v , and let the horizontal distance for removing overlap be d_h . FT moves the overlapping node horizontally if $d_h < d_v$, else vertically, and experimentally, it has been shown that FT gives a layout of smaller area than FS and FS’.

FS, FS’ and FT belong to the family of force based layout algorithms. Spring based algorithms treat edges as springs obeying Hooke’s Law, and the nodes are pushed apart or pulled in iteratively to balance the forces till an equilibrium is reached. A spring based algorithm *ODNLS*, which adjusts the attractive/repulsive force between two nodes dynamically, is proposed in [16], which preserves the orthogonal ordering of the input layout and typically returns a smaller overlap-free layout than the force-based family of algorithms.

It is worth noting that none of the algorithms mentioned above give a provable worst-case guarantee on the quality of the output.

The hardness of preserving orthogonal constraints w.r.t various optimization metrics has also been well-studied. Brandes and Pampel [3] showed that it is NP-hard to determine if there exists an orthogonal-order preserving rectilinear drawing of a simple path, and extend the result for determination of uniform edge-length drawings of simple paths with same constraints. LADR was shown to be NP-hard by Hayashi et al. [12], using a reduction from 3SAT.

1.2 Related Work

Algorithms for label placement and packing that do not account for orthogonal ordering have been extensively studied. The placement of labels corresponding to points on a map is a natural problem that arises in geographic information systems (GIS) [22]. In particular, placing labels on maps such that the label boundary coincides with the point feature has been a well-studied problem. A common objective in such label-placement problems is to maximize the number of features labelled, such that the labels are pairwise disjoint. We refer to [2, 15] as examples of this line of work.

Packing rectangles without orthogonality constraints has also been well-studied. One such problem is the strip packing problem, in which we want to pack a set of rectangles into a strip of given width while minimizing the height of the packing. It is known that the strip-packing problem is strongly NP-hard [17]. It can be easily seen that if the constraint for orthogonal order preservation is removed, then LADR can be reduced to multiple instances of strip packing problem. There has been extensive work done on strip packing [20, 19, 10], with the current best algorithm being a $5/3 + \varepsilon$ -approximation by Harren et al. [9].

Another related packing problem is the two-dimensional geometric knapsack problem, defined as follows. The input consists of a set of weighted rectangles and a rectangular knapsack, and the goal is to find a subset of rectangles of maximum weight that can be placed in the knapsack such that no two rectangles have an overlap. The 2D-knapsack problem is known to be strongly NP-hard even when the input consists of a set of unweighted squares [17]. Recently, Adamaszek and Wiese [1] gave a quasi-polynomial time $(1 + \varepsilon)$ approximation scheme for this problem, with the assumption that the input consists of quasi-polynomially bounded integers.

1.3 Our results

We point out an intimate connection between LADR and the problem of hitting segments using a minimum number of horizontal and vertical lines. In particular, the segments to be hit are the ones connecting each pair of rectangle centers in the input layout. The connection to the hitting set is described in [Section 3](#). To our knowledge, this connection to hitting sets has not been observed in the literature. We exploit the connection to hitting set to prove hardness results for LADR in [Section 4](#) that complement the NP-completeness result in [12]. We show that it is APX-hard to find a layout that minimizes the perimeter of the bounding box. We also show that if there is an approximate decision procedure that determines whether there is a layout that fits within a bounding box of specified dimensions, then $\mathbb{P} = \text{NP}$. These hardness results hold even when the input rectangles are unit squares. The results for LADR follow from a hardness of approximation result that we show for a hitting set problem. The starting point of the latter is the result of Hassin and Megiddo [11] who show that it is NP-hard to determine if there is a set of k axis-parallel lines that hit a set of horizontal segments of unit length. The added difficulty that we need to overcome is that in our case, the set of segments that need to be hit cannot be

arbitrarily constructed. Rather, the set consists of all segments induced by a set of arbitrarily constructed points.

It is possible to exploit this connection to hitting sets and use known algorithms for hitting sets (e.g. [8]) to devise an $O(1)$ approximation algorithm for LADR. Instead, we describe (in Section 5) a direct polynomial time algorithm for LADR that achieves a $4(1+o(1))$ approximation. This is the first polynomial time algorithm for LADR with a provable approximation guarantee. The algorithm involves solving a linear-programming relaxation of LADR followed by a simple rounding.

2 Preliminaries

We define a layout λ of a set of rectangles R as an assignment of coordinates to the center of each rectangle $r \in R$ i.e. $\lambda : R \rightarrow \mathbb{R}^2$. Our input for LADR consists of a set of rectangles R , and an initial layout λ^{in} . We will assume that λ^{in} is injective, i.e. no two rectangle centers coincide in the input layout. A rectangle r is defined by its horizontal width w_r and vertical height h_r , both of which are assumed to be integral. It is given that all rectangles are axis-parallel in λ^{in} , and rotation of rectangles is not allowed in any adjusted layout.

The coordinates of center of r in layout λ is denoted by $\lambda(r) = (x_r, y_r)$. For brevity, we denote the x -coordinate of $\lambda(r)$ by $\lambda_x(r)$, and the corresponding y -coordinate by $\lambda_y(r)$. The set of points $\{\lambda(r) : r \in R\}$ is denoted by $\lambda(R)$.

A pair of rectangles $r, r' \in R$ is said to intersect in a layout λ if and only if

$$|\lambda_x(r) - \lambda_x(r')| < \frac{w_r + w_{r'}}{2} \quad \text{and} \quad |\lambda_y(r) - \lambda_y(r')| < \frac{h_r + h_{r'}}{2}. \quad (1)$$

A layout λ is termed as a *disjoint* layout if no two rectangles in R intersect with each other. Let $W_l(\lambda)$ and $W_r(\lambda)$ denote the x -coordinates of the left and right sides of the smallest axis-parallel rectangle bounding the rectangles of R placed by λ , respectively. We then define the width of the layout, $W(\lambda) = W_r(\lambda) - W_l(\lambda)$. Similarly, let $H_t(\lambda)$ and $H_b(\lambda)$ define the y -coordinates of the top and bottom of the bounding rectangle, and the height of the layout is defined as $H(\lambda) = H_t(\lambda) - H_b(\lambda)$. The area of λ is thus defined as $A(\lambda) = H(\lambda) \times W(\lambda)$. The perimeter of λ is $2(H(\lambda) + W(\lambda))$.

Let λ and λ' be two layouts of R . Then, λ and λ' are defined to have the same *orthogonal ordering* if for any two rectangles $r, r' \in R$,

$$\lambda_x(r) < \lambda_x(r') \iff \lambda'_x(r) < \lambda'_x(r') \quad (2)$$

$$\lambda_y(r) < \lambda_y(r') \iff \lambda'_y(r) < \lambda'_y(r') \quad (3)$$

$$\lambda_x(r) = \lambda_x(r') \iff \lambda'_x(r) = \lambda'_x(r') \quad (4)$$

$$\lambda_y(r) = \lambda_y(r') \iff \lambda'_y(r) = \lambda'_y(r') \quad (5)$$

For any R and corresponding λ^{in} , the minimal area of a layout is defined as: $A^{\text{min}} = \inf\{A(\lambda) : \lambda \text{ is a disjoint layout, } \lambda \text{ has same orthogonal ordering as } \lambda^{\text{in}}\}$. It should be noted that it may not be possible to attain a disjoint orthogonality

preserving layout whose area is the same as A^{\min} - we can only aim to get a layout whose area is arbitrarily close to A^{\min} .

We introduce the concept of rigidities in a layout. A set of rectangles $R' \subseteq R$ forms a x -rigidity in a layout λ if $\exists \alpha$ such that $R' = \{r \in R \mid \lambda_x(r) = \alpha\}$. We define a y -rigidity in a layout similarly in terms of the y -coordinates of the rectangles in that layout. We observe that any rectangle r belongs to a unique x -rigidity (and a unique y -rigidity), which may consist of merely itself and no other rectangle. We order the x -rigidities in a layout λ in increasing order of x -coordinates, and for any rectangle $r \in R$, we define its x -rank to be i if r belongs to the i -th x -rigidity in this ordering. It is obvious that x -rank of any rectangle is an integer between 1 and $|R|$. We similarly define the y -rank of each rectangle in terms of its y -rigidities. Unless otherwise stated, we refer to the rigidities and ranks of the initial layout λ^{in} whenever these terms are used in the paper.

Let $\phi(p, p')$ be the segment whose endpoints are points $p, p' \in P$. Then the set of segments induced by a set of points P is defined as $\Phi(P) = \{\phi(p, p') : p, p' \in P, p \neq p'\}$, denoted by Φ when P is clear from the context.

We also consider a simpler version of LADR where the set of rectangles R consists of unit squares. We call this version as the *Layout Adjustment for Disjoint Squares* problem, and refer to it as LADS for brevity.

3 Reduction of LADS to Hitting Set

We formally define a unit grid as follows. Let $f : \mathbb{R}^2 \rightarrow \mathbb{Z}^2$ be the function $f(x, y) = (\lfloor x \rfloor, \lfloor y \rfloor)$. The function f induces a partition of \mathbb{R}^2 into grid cells - grid cell (i, j) is the set $\{p \in \mathbb{R}^2 \mid f(p) = (i, j)\}$. We call this partition a unit grid on \mathbb{R}^2 . The ‘grid lines’ are the vertical lines $x = \alpha$ and $y = \alpha$ for integer α .

Let S be the set of unit squares provided as input to LADS, having initial layout λ^{in} . Consider a disjoint, orthogonal order preserving layout λ for S . Let L be the subset consisting of those grid lines that intersect the minimum bounding box of $\lambda(S)$. Let ϕ be the line segment connecting the points $\lambda(s)$ and $\lambda(s')$, for some $s, s' \in S$. Since the layout λ is disjoint, $\lambda(s)$ and $\lambda(s')$ lie in different grid cells. Thus, there exists at least one line $\tau \in L$ that intersects ϕ . Motivated by this, we define a hitting set problem as follows.

We say a line τ hits a line segment ϕ if τ intersects the relative interior of ϕ but not either end point of ϕ . Thus, if ϕ is a horizontal line segment (which would happen if s, s' belongs to a y -rigidity), then ϕ cannot be hit by a horizontal line $\tau \in L$. We thus define the Uniform Hitting Set (UHS) problem as follows:

Definition 1 (Uniform Hitting Set - Decision Problem). Given a set of segments Φ induced by a point set P and a non-negative integer k , is there a set of axis-parallel lines L that hit all segments in Φ , such that $|L| \leq k$?

Since the area of the minimum bounding box for $\lambda(S)$ is roughly the product of the number of horizontal grid lines intersecting it and the number of vertical grid lines intersecting it, we also need the following variant.

Definition 2 (Constrained Uniform Hitting Set - Decision Problem). Given a set of line segments, Φ , induced by a set of points P , and non negative

integers r, c , is it possible to hit all segments in Φ with a set of lines L containing at most r horizontal lines and c vertical lines ?

The term ‘uniform’ in the problem name refers to the fact that each segment in Φ needs to be hit only once by a horizontal or vertical line. We denote the problem thus defined as CUHS, and proceed to show its equivalence with a constrained version of the layout adjustment problem.

Definition 3 (Constrained LADS - Decision Problem). Given n unit squares S , initial layout λ^{in} , positive integers w, h and a constant $0 < \varepsilon < 1$, is there a layout λ' having height $H(\lambda') \leq h + \varepsilon$ and width $W(\lambda') \leq w + \varepsilon$, satisfying the following conditions?

1. λ' is a disjoint layout.
2. λ^{in} and λ' have the same orthogonal order.

We term the constrained version of layout adjustment problem as CLADS. We now show how to reduce a given instance of CLADS into an instance of CUHS. We define Φ as the set of all line segments induced by points in $\lambda^{\text{in}}(S)$.

Lemma 1. *If there is a set of lines L containing at most r horizontal lines and at most c vertical lines that hit all segments in Φ , then there is a disjoint layout λ' that has the same orthogonality as λ^{in} and whose height and width is bounded by $h + \varepsilon$ and $w + \varepsilon$, for any $\varepsilon > 0$. Here $h = r + 1$, $w = c + 1$.*

To solve LADS by multiple iterations of a procedure for solving CUHS, it would be useful to guess the width of a disjoint layout with near-optimal area. The following observation allows us to restrict our attention to layouts with near integral width. That makes it possible to discretize LADS, by solving a constrained version of LADS for all values of widths in $\{1, 2, \dots, |S|\}$.

Lemma 2. *Any disjoint layout λ can be modified into a disjoint layout λ' having the same height and orthogonal ordering as λ , such that $W(\lambda') (\leq W(\lambda))$ lies in the interval $[w, w + \varepsilon]$, where $w \in \{1, 2, \dots, n\}$ and $\varepsilon > 0$ is an arbitrarily small constant.*

We can similarly modify a disjoint layout λ into an orthogonal order preserving disjoint layout λ' which has the same width, and whose height lies in the interval $[h, h + \varepsilon]$ for some integer $h > 0$. Thus, combining the two methods, we obtain the following corollary:

Corollary 3. *Any disjoint layout λ can be modified into an orthogonal order preserving disjoint layout λ' , such that $W(\lambda') (\leq W(\lambda))$ lies in the interval $[w, w + \varepsilon]$ and $H(\lambda') (\leq H(\lambda))$ lies in the interval $[h, h + \varepsilon]$, where $w, h \in \{1, 2, \dots, n\}$ and $\varepsilon > 0$ is an arbitrarily small constant.*

Lemma 4. *For any $\varepsilon < 1/2$, if there is a disjoint layout λ' that has the same orthogonality as λ^{in} and whose height and width is bounded by $h + \varepsilon$ and $w + \varepsilon$ respectively, where h, w are positive integers, then there is a set of lines L that hit all segments in Φ , containing at most c vertical lines and r horizontal lines. Here $r = h - 1$, $c = w - 1$.*

All proofs of lemmas in this section are in [Appendix A.1](#). [Lemmas 1](#) and [4](#) and [Corollary 3](#) show the close connection between CLADS and CUHS. In subsequent sections, we exploit this connection to derive hardness results for CLADS.

4 Inapproximability of Layout Adjustment Problems

In this section, we prove APX-hardness of various layout adjustment problems. We consider a variant of LADS where instead of minimizing the area, we would like to minimize the perimeter of the output layout. We prove an inapproximability result for this problem which readily follows from APX-hardness of the Uniform Hitting Set problem. We also show that the decision problem Constrained LADS (CLADS) is NP-hard. Recall that in this problem, given an initial layout of n unit squares, positive integers w, h , and a constant $\varepsilon > 0$, the goal is to determine if there is an orthogonal order preserving layout having height and width at most $h + \varepsilon$ and $w + \varepsilon$ respectively. To be precise we show a more general inapproximability result for this problem. We prove that, given an instance of CLADS, it is NP-hard to determine whether there is an output layout of height and width at most $h + \varepsilon$ and $w + \varepsilon$ respectively, or there is no output layout of respective height and width at most $(1 + \xi)(h + \varepsilon)$ and $(1 + \xi)(w + \varepsilon)$ for some $0 < \xi < 1$. This result follows from the connection of CLADS with Constrained Uniform Hitting Set (CUHS) described in [Section 3](#) and APX-hardness of CUHS. The APX-hardness of CUHS follows from the APX-hardness of UHS, to which we turn next.

APX-Hardness of Hitting Set Problem. We consider the optimization version of UHS, in which given a set of points P , the goal is to find minimum number of vertical and horizontal lines that hit all segments in $\Phi(P)$. In this section, we prove that there is no polynomial time $(1 + \xi)$ -factor approximation algorithm for UHS, unless $\mathbb{P} = \text{NP}$, for some $0 < \xi < 1$. Note that the UHS problem we consider here is a special case of the hitting set problem where, given any set of segments S , the goal is to find a hitting set for S . This problem is known to be NP-hard. But, in case of UHS, given a set of points, we need to hit all the segments induced by the points. Thus the nontriviality in our result is to show that even this special case of hitting set is not only NP-hard, but also hard to approximate. To prove the result we reduce a version of maximum satisfiability problem (5-OCC-MAX-3SAT) to UHS. 5-OCC-MAX-3SAT is defined as follows. Given a set X of n boolean variables and a conjunction ϕ of m clauses such that each clause contains precisely three distinct literals and each variable is contained in exactly five clauses ($m = \frac{5n}{3}$), the goal is to find a binary assignment of the variables in X so that the maximum number of clauses of ϕ are satisfied. The following theorem follows from the work of Feige [7].

Theorem 5. *For some $\gamma > 0$, it is NP-hard to distinguish between an instance of 5-OCC-MAX-3SAT consisting of all satisfiable clauses, and one in which less than $(1 - \gamma)$ -fraction of the clauses can be satisfied.*

The crux of the hardness result is to show the existence of a reduction from 5-OCC-MAX-3SAT to UHS having the following properties:

1. Any instance of 5-OCC-MAX-3SAT in which all the clauses can be satisfied, is reduced to an instance of UHS in which the line segments in $\Phi(P)$ can be hit using at most k lines, where k is a function of m and n .
2. Any instance of 5-OCC-MAX-3SAT in which less than $1 - \delta$ (for $0 < \delta \leq 1$) fraction of the clauses can be satisfied, is reduced to an instance of UHS in which more than $(1 + \frac{1}{55}\delta)k$ lines are needed to hit the segments in $\Phi(P)$.

The complete reduction appears in [Appendix A.3](#). The next theorem follows from the existence of such a reduction and from [Theorem 5](#).

Theorem 6. *There is no polynomial time $(1 + \xi)$ -factor approximation algorithm for UHS with $\xi \leq \frac{1}{55}\gamma$, unless $\mathbb{P} = \mathbb{NP}$, γ being the constant in [Theorem 5](#).*

Now we consider the variant of LADS where we would like to minimize the perimeter $2(w + v)$ of the output layout, where w and v are the width and height of the layout respectively. We refer to this problem as Layout Adjustment for Disjoint Squares - Minimum Perimeter (LADS-MP). We note that in UHS we minimize the sum of the number of horizontal and vertical lines ($k = r + c$). Thus by [Lemma 1](#) and [Lemma 4](#) it follows that a solution for UHS gives a solution for LADS-MP (within an additive constant) and vice versa. Hence the following theorem easily follows from [Theorem 6](#).

Theorem 7. *No polynomial time $(1 + \xi')$ -factor approximation algorithm exists for LADS-MP with $\xi' = \frac{\xi}{4}$, unless $\mathbb{P} = \mathbb{NP}$, ξ being the constant in [Theorem 6](#).*

Inapproximability of CUHS. We show that if there is a polynomial time approximate decision algorithm for Constrained Uniform Hitting Set - Decision Problem (CUHS), then $\mathbb{P} = \mathbb{NP}$. We use the inapproximability result of UHS for this purpose. See [Definition 2](#) for the definition of CUHS. Now we have the following theorem whose proof follows from [Theorem 6](#) and is given in [Appendix A.3.1](#).

Theorem 8. *Suppose there is a polynomial time algorithm that, given $\Phi(P)$ and non-negative integers r, c as input to CUHS,*

- (1) *outputs “yes”, if there is a set with at most c vertical and r horizontal lines that hits the segments in $\Phi(P)$; and*
- (2) *outputs “no”, if there is no hitting set for $\Phi(P)$ using at most $(1 + \xi)c$ vertical and $(1 + \xi)r$ horizontal lines, where ξ is the constant in [Theorem 6](#).*

Then $\mathbb{P} = \mathbb{NP}$.

Inapproximability of CLADS. We show that the existence of a polynomial time approximate decision algorithm for CLADS implies $\mathbb{P} = \mathbb{NP}$. See [Definition 3](#) for the definition of CLADS. Now we have the following theorem whose proof follows from [Theorem 8](#) and is given in [Appendix A.3.2](#).

Theorem 9. Suppose there is a polynomial time algorithm that, given S , λ^{in} , w, h , and ε as input to CLADS,

- (1) outputs “yes”, if there is an output layout λ' with $H(\lambda') \leq h + \varepsilon$ and $W(\lambda') \leq w + \varepsilon$; and
- (2) outputs “no”, if there is no output layout λ' with $H(\lambda') \leq (1 + \xi')(h + \varepsilon)$ and $W(\lambda') \leq (1 + \xi')(w + \varepsilon)$, where $\xi' = \frac{\xi}{4}$ and ξ is the constant in [Theorem 6](#).

Then $\mathbb{P} = \mathbb{NP}$.

5 Approximation Algorithm

In this section, we describe an approximation algorithm for LADR i.e. for a set R of axis-parallel rectangles having initial layout λ^{in} , we need to find a disjoint layout of minimum area that preserves the orthogonal ordering of λ^{in} . Let $W_{\max} = \max\{w_r \mid r \in R\}$ and $H_{\max} = \max\{h_r \mid r \in R\}$ be the maximum width and maximum height, respectively, amongst all rectangles in R . [Lemma 2](#) showed that if the input consists of a set of squares S , any disjoint layout of S can be modified into a disjoint layout having same orthogonality such that its width is arbitrarily close to an integer from the set $\{1, \dots, |S|\}$. It can be seen that [Lemma 2](#) can be extended in a straightforward manner for a set of axis-parallel rectangles R i.e. any disjoint layout of R can be modified into a disjoint orthogonal-order preserving layout having a width that is arbitrarily close to an integer from the set $\{W_{\max}, W_{\max} + 1, \dots, W_R\}$, where $W_R = \sum_{r \in R} w_r$. We henceforth state [Corollary 3](#) in the context of LADR as follows.

Corollary 10. Let $W_R = \sum_{r \in R} w_r$ and $H_R = \sum_{r \in R} h_r$ be the sum of the widths and sum of the heights of all the rectangles in R , respectively. Then, any disjoint layout λ of R can be modified into an orthogonal order preserving layout λ' of R , such that $W(\lambda') (\leq W(\lambda))$ lies in the interval $[w, w + \varepsilon]$ and $H(\lambda') (\leq H(\lambda))$ lies in the interval $[h, h + \varepsilon]$, where $w \in \{W_{\max}, W_{\max} + 1, \dots, W_R\}$, $h \in \{H_{\max}, H_{\max} + 1, \dots, H_R\}$ and $\varepsilon > 0$ is an arbitrarily small constant.

Using [Corollary 10](#), we know that for any disjoint layout λ of R , there is a corresponding disjoint layout λ' having the same orthogonal order as λ , whose height and width are arbitrarily close to an integer from a known set of integers. Hence, we look at all disjoint orthogonality preserving layouts in that range, and choose the one with the minimum area as our solution.

Given positive integers $w \in \{W_{\max}, W_{\max} + 1, \dots, W_R\}$, $h \in \{H_{\max}, H_{\max} + 1, \dots, H_R\}$ and $0 < \varepsilon < 1$, we formulate as a LP the problem of whether there is an orthogonal order preserving layout λ with $W(\lambda) \leq w + \varepsilon$, $H(\lambda) \leq h + \varepsilon$. Recall that a layout λ assigns a location $\lambda(r) = (x_r, y_r)$ for the center of each rectangle $r \in R$. The variables of our linear program are $\cup_{r \in R} \{x_r, y_r\}$. For any two rectangles $r, r' \in R$, $\lambda_x^{\text{in}}(r) < \lambda_x^{\text{in}}(r')$ implies that $x_r < x'_r$. We add such a constraint for each pair of rectangles in R , both for x -coordinate and y -coordinate of the layout. Similarly, we add the constraint $x_r = x'_r$ for all pair of rectangles

$r, r' \in R$ for which $\lambda_x^{\text{in}}(r) = \lambda_x^{\text{in}}(r')$. These constraints ensure orthogonality is preserved in the output layout.

We now look at constraints that ensure disjointness of the output layout. Let r and r' be two rectangles in the initial layout λ^{in} , having dimensions (w_r, h_r) and $(w_{r'}, h_{r'})$ respectively. We define $w(r, r') = \frac{w_r + w_{r'}}{2}$ and $h(r, r') = \frac{h_r + h_{r'}}{2}$. Let $x_{\text{diff}}(r, r') = \begin{cases} x_r - x_{r'}, & \text{if } \lambda_x^{\text{in}}(r') \leq \lambda_x^{\text{in}}(r) \\ x_{r'} - x_r, & \text{otherwise} \end{cases}$. We define $y_{\text{diff}}(r, r')$ analogously.

If r, r' are disjoint in some layout, then either their x -projections or their y -projections are disjoint in that layout. Equivalently, either the difference in x -coordinates of the centers of rectangles r, r' is at least $w(r, r')$, or the difference in y -coordinates of the centers is at least $h(r, r')$. We thus get the following LP:

$$x_r < x_{r'} \quad \forall r, r' \in R : \lambda_x^{\text{in}}(r) < \lambda_x^{\text{in}}(r') \quad (6)$$

$$x_r = x_{r'} \quad \forall r, r' \in R : \lambda_x^{\text{in}}(r) = \lambda_x^{\text{in}}(r') \quad (7)$$

$$y_r < y_{r'} \quad \forall r, r' \in R : \lambda_y^{\text{in}}(r) < \lambda_y^{\text{in}}(r') \quad (8)$$

$$y_r = y_{r'} \quad \forall r, r' \in R : \lambda_y^{\text{in}}(r) = \lambda_y^{\text{in}}(r') \quad (9)$$

$$\left(x_{r'} + \frac{w_{r'}}{2}\right) - \left(x_r - \frac{w_r}{2}\right) \leq w + \varepsilon \quad \forall r, r' \in R : \lambda_x^{\text{in}}(r) < \lambda_x^{\text{in}}(r') \quad (10)$$

$$\left(y_{r'} + \frac{h_{r'}}{2}\right) - \left(y_r - \frac{h_r}{2}\right) \leq h + \varepsilon \quad \forall r, r' \in R : \lambda_y^{\text{in}}(r) < \lambda_y^{\text{in}}(r') \quad (11)$$

$$\frac{x_{\text{diff}}(r, r')}{w(r, r')} + \frac{y_{\text{diff}}(r, r')}{h(r, r')} \geq 1 \quad \forall r, r' \in R \quad (12)$$

Inequalities (6) to (9) model the orthogonal ordering requirement for a layout, while **Inequalities (10) to (11)** restrict the width and height of the layout respectively. Since any two rectangles r, r' in a disjoint layout are separated by at least half the sum of their widths in the x -direction ($w(r, r')$) or at least half the sum of their heights in the y -direction ($h(r, r')$), **Inequality (12)** ensures that every such layout is a valid solution for the linear program. We incorporate the linear program into **Algorithm 1** for solving LADR.

Algorithm 1 *ApproxLADR*(R, λ^{in})

Input: A set of rectangles R , and initial layout λ^{in} .

Output: A disjoint layout that has the same orthogonal order as λ^{in} .

- 1: **for** $w = W_{\text{max}}$ to W_R **do**
 - 2: **for** $h = H_{\text{max}}$ to H_R **do**
 - 3: **if** LP stated in **Inequalities (6) to (12)** is feasible **then**
 - 4: $\lambda_{w,h} \leftarrow$ Layout returned by solution of LP.
 - 5: **if** λ^{min} is undefined **or** $A(\lambda_{w,h}) < A(\lambda^{\text{min}})$ **then**
 - 6: $\lambda^{\text{min}} \leftarrow \lambda_{w,h}$
 - 7: Define $\lambda(R) = 2 \cdot \lambda^{\text{min}}(R)$ i.e. $\lambda(r) = (2 * \lambda_x^{\text{min}}(r), 2 * \lambda_y^{\text{min}}(r))$, $\forall r \in R$
 - 8: **return** The layout λ .
-

Lemma 11. *ApproxLADR(R, λ^{in}) returns a 4-approximation for LADR.*

Proof. Let $\lambda_{w,h}$ be any feasible layout returned by the LP in Line 4, for some value of w, h . Let r, r' be two rectangles in R , and assume that $\lambda_x^{\text{in}}(r) > \lambda_x^{\text{in}}(r')$, $\lambda_y^{\text{in}}(r) > \lambda_y^{\text{in}}(r')$. (The other cases are symmetric). By Inequality (12), either $\frac{x_{\text{diff}}(r, r')}{w(r, r')} \geq \frac{1}{2}$ or $\frac{y_{\text{diff}}(r, r')}{h(r, r')} \geq \frac{1}{2}$. Without loss of generality, assume its the former. Consider the layout $\lambda = 2\lambda_{w,h}$, as in Line 7. Hence, our assumption that $\frac{x_{\text{diff}}(r, r')}{w(r, r')} \geq \frac{1}{2}$ implies that $\lambda_x(r) - \lambda_x(r') = 2x_r - 2x_{r'} \geq w(r, r')$, which satisfies the criteria for disjointness in Inequality (1). Since the final layout λ returned by the algorithm equals $2 \cdot \lambda_{w',h'}$ for some feasible layout $\lambda_{w',h'}$, λ is a disjoint layout that also satisfies the constraints for orthogonal ordering in Inequalities (6) to (9).

Let λ^* be any disjoint layout preserving the orthogonal ordering of λ^{in} . We may assume, by Corollary 10, that its width is at most $w' + \varepsilon$ and its height is at most $h' + \varepsilon$, for some integers $w' \in \{W_{\max}, W_{\max} + 1, \dots, W_R\}$, $h' \in \{H_{\max}, H_{\max} + 1, \dots, H_R\}$ and some $\varepsilon > 0$. Consider the iteration of the inner for loop in Algorithm 1 with $w = w'$ and $h = h'$. Since λ^* is a valid solution for the LP, the layout $\lambda_{w,h}$ computed in Line 4 (and hence λ^{min}) has an area that is less than or equal to that of λ^* . The algorithm Algorithm 1 returns a layout $\lambda(S)$ obtained by multiplying each of the coordinates in λ^{min} by a factor of 2. Hence, the layout $\lambda(S)$ has at most twice the width and at most twice the height of λ^* , ensuring that $A(\lambda) \leq 4 * A(\lambda^*)$. \square

We note that since W_R, H_R are not polynomial in the input size, the resultant algorithm is a pseudo-polynomial time algorithm. But by searching across exponentially increasing value of widths, and thereby losing a small approximation factor, we can obtain a $4(1 + o(1))$ polynomial time approximation for LADR. We conclude by summarizing our result as follows:

Theorem 12. *There is a polynomial time algorithm that returns a $4(1 + o(1))$ -approximation for LADR i.e. given a set of rectangles R and an initial layout λ^{in} , it returns an orthogonal order preserving disjoint layout whose area is at most $4(1 + o(1))$ times the area attainable by any such layout.*

We note that our approach can also be used to get a $2(1 + o(1))$ approximation for the problem of finding a layout of rectangles that minimizes the perimeter.

References

- [1] Adamaszek, A., Wiese, A.: A quasi-ptas for the two-dimensional geometric knapsack problem. In: SODA. pp. 1491–1505 (2015)
- [2] Agarwal, P.K., van Kreveld, M.J., Suri, S.: Label placement by maximum independent set in rectangles. Comput. Geom. 11(3-4), 209–218 (1998)
- [3] Brandes, U., Pampel, B.: Orthogonal-ordering constraints are tough. J. Graph Algorithms Appl. 17(1), 1–10 (2013)

- [4] Bridgeman, S., Tamassia, R.: Difference metrics for interactive orthogonal graph drawing algorithms. In: Graph Drawing, pp. 57–71 (1998)
- [5] Bridgeman, S., Tamassia, R.: A user study in similarity measures for graph drawing. *J. Graph Algorithms Appl.* 6(3), 225–254 (2002)
- [6] Eades, P., Lai, W., Misue, K., Sugiyama, K.: Preserving the mental map of a diagram. International Institute for Advanced Study of Social Information Science, Fujitsu Limited (1991)
- [7] Feige, U.: A threshold of $\ln n$ for approximating set cover. *J. ACM* 45(4), 634–652 (1998)
- [8] Gaur, D.R., Ibaraki, T., Krishnamurti, R.: Constant ratio approximation algorithms for the rectangle stabbing problem and the rectilinear partitioning problem. *J. Algorithms* 43(1), 138–152 (2002)
- [9] Harren, R., Jansen, K., Prädél, L., van Stee, R.: A $(5/3 + \varepsilon)$ -approximation for strip packing. *Comput. Geom.* 47(2), 248–267 (2014)
- [10] Harren, R., van Stee, R.: Improved absolute approximation ratios for two-dimensional packing problems. In: APPROX-RANDOM. pp. 177–189 (2009)
- [11] Hassin, R., Megiddo, N.: Approximation algorithms for hitting objects with straight lines. *Discrete Appl. Math.* 30(1), 29–42 (Jan 1991)
- [12] Hayashi, K., Inoue, M., Masuzawa, T., Fujiwara, H.: A layout adjustment problem for disjoint rectangles preserving orthogonal order. In: Graph Drawing. pp. 183–197 (1998)
- [13] Herman, I., Melançon, G., Marshall, M.S.: Graph visualization and navigation in information visualization: A survey. *Visualization and Computer Graphics*, IEEE Transactions on 6(1), 24–43 (2000)
- [14] Huang, X., Lai, W., Sajeew, A., Gao, J.: A new algorithm for removing node overlapping in graph visualization. *Information Sciences* 177(14), 2821–2844 (2007)
- [15] van Kreveld, M.J., Strijk, T., Wolff, A.: Point labeling with sliding labels. *Comput. Geom.* 13(1), 21–47 (1999)
- [16] Li, W., Eades, P., Nikolov, N.S.: Using spring algorithms to remove node overlapping. In: APVIS. CRPIT, vol. 45, pp. 131–140 (2005)
- [17] Lodi, A., Martello, S., Monaci, M.: Two-dimensional packing problems: A survey. *European Journal of Operational Research* 141(2), 241–252 (2002)
- [18] Misue, K., Eades, P., Lai, W., Sugiyama, K.: Layout adjustment and the mental map. *Journal of visual languages and computing* 6(2), 183–210 (1995)
- [19] Schiermeyer, I.: Reverse-fit: A 2-optimal algorithm for packing rectangles. In: ESA. pp. 290–299 (1994)
- [20] Steinberg, A.: A strip-packing algorithm with absolute performance bound 2. *SIAM J. Comput.* 26(2), 401–409 (1997)
- [21] Storey, M.A.D., Müller, H.A.: Graph layout adjustment strategies. In: Graph Drawing. pp. 487–499 (1996)
- [22] Wolff, A.: The Map-Labeling bibliography. http://i11www.iti.uni-karlsruhe.de/~awolff/map-labeling/bibliography/maplab_date.html (2009), accessed: 2014-08-14

A Appendix

A.1 Proofs for Section 3

Lemma 1. *If there is a set of lines L containing at most r horizontal lines and at most c vertical lines that hit all segments in Φ , then there is a disjoint layout λ' that has the same orthogonality as λ^{in} and whose height and width is bounded by $h + \varepsilon$ and $w + \varepsilon$, for any $\varepsilon > 0$. Here $h = r + 1$, $w = c + 1$.*

Proof. For any $p \in P$, we define $\rho(p) : P \rightarrow \mathbb{Z}^2$ as $\rho(p) = (i, j)$ where i is the number of vertical lines in L to the left of p , and j is the number of horizontal lines in L below p . We note that the function $\rho : P \rightarrow \mathbb{Z}^2$ is one-to-one, as otherwise, the segment corresponding to two points p and p' with $\rho(p) = \rho(p')$ would not be hit by any line in L . For convenience, if $\rho(p) = (i, j)$, we denote $\rho_x(p) = i$ and $\rho_y(p) = j$.

We first consider the placement $\lambda(p) = (i + \frac{1}{2}, j + \frac{1}{2})$, where $i = \rho_x(p)$ and $j = \rho_y(p)$. This ensures disjointness, but not orthogonality - within a single column (or row) in λ , there could be orthogonality violations w.r.t λ^{in} due to all endpoints having the same x (or y) coordinate. This is fixed as follows.

We define $P_x(i) = \{p \in P \mid \rho_x(p) = i\}$, which represents the i -th column of the hitting set. We similarly define $P_y(j) = \{p \in P \mid \rho_y(p) = j\}$. Let $\alpha_i(p)$ denote the x -rank of p within $P_x(i)$, and similarly let $\beta_j(p)$ denote the y -rank of p within $P_y(j)$. For some $\delta > 0$, we define the layout λ' for a point p with $\rho(p) = (i, j)$ as:

$$\lambda'(p) = \left(i(1 + \delta) + \frac{1}{2} + \delta \cdot \frac{\alpha_i(p)}{n}, j(1 + \delta) + \frac{1}{2} + \delta \cdot \frac{\beta_j(p)}{n} \right) \quad (13)$$

We observe that the maximum value of $\alpha_i(p)$ (or $\beta_j(p)$) is n . Thus, for any 2 points p, p' in consecutive columns i.e. $p \in P_x(i), p' \in P_x(i+1)$, $|\lambda'_x(p) - \lambda'_x(p')| \geq 1$, making the corresponding unit squares disjoint. Any 2 points in consecutive rows are similarly placed more than unit distance apart, which establishes the disjointness of the layout.

We now show that λ^{in} and λ' have the same orthogonal ordering. Let $p, p' \in P$ be two unit squares and assume $\lambda_x^{\text{in}}(p) \leq \lambda_x^{\text{in}}(p')$. If $\rho_x(p) < \rho_x(p')$, it is clear that $\lambda'_x(p) < \lambda'_x(p')$ as desired. If $\rho_x(p) = \rho_x(p') = i$, then it is easy to see that $\lambda_x^{\text{in}}(p) < \lambda_x^{\text{in}}(p') \Rightarrow \alpha_i(p) < \alpha_i(p') \Rightarrow \lambda'_x(p) < \lambda'_x(p')$ and $\lambda_x^{\text{in}}(p) = \lambda_x^{\text{in}}(p') \Rightarrow \alpha_i(p) = \alpha_i(p') \Rightarrow \lambda'_x(p) = \lambda'_x(p')$. Similar reasoning applies to the y -coordinates.

Since L has at most c vertical lines ($c \leq n - 1$), for any $p, p' \in P$, $|\lambda'_x(p) - \lambda'_x(p')| \leq c(1 + \delta) + \delta = c + (c + 1)\delta$. Since this is the maximum difference of x -coordinates between any two centers of unit squares in the layout, $W(\lambda') \leq c + 1 + (c + 1)\delta = w + (c + 1)\delta$. Putting the value of $\delta = \frac{\varepsilon}{n}$, we have $W(\lambda') \leq w + \varepsilon$. Similarly, we can show that $H(\lambda') \leq h + \varepsilon$, thus concluding the proof. \square

Lemma 2. *Any disjoint layout λ can be modified into a disjoint layout λ' having the same height and orthogonal ordering as λ , such that $W(\lambda') (\leq W(\lambda))$ lies in the interval $[w, w + \varepsilon]$, where $w \in \{1, 2, \dots, n\}$ and $\varepsilon > 0$ is an arbitrarily small constant.*

Proof. Let s_i denote the set of squares in λ with x -rank i , and let $\delta > 0$ be a parameter. We create a modified layout λ' by “compressing” the layout λ horizontally while keeping the y -coordinates untouched. We set $\lambda'_x(s_1) = 1$, and assign the x -coordinates of the remaining squares in ascending order of their x -ranks in λ . We claim that it is possible to place squares in s_i such that $\lambda'_x(s_i) \in [b, b + (i - 1) \cdot \delta]$, where $0 \leq b \leq i - 1 \in \mathbb{Z}_+ \cup \{0\}$.

For the base case, we place the squares in s_2 . If the projection of s_2 on the y -axis is not disjoint from that of s_1 , then set $\lambda'_x(s_2) = \lambda'_x(s_1) + 1$. If not, set $\lambda'_x(s_2) = \lambda'_x(s_1) + \delta$. In both cases, s_1 and s_2 do not satisfy **Inequality (1)** in layout λ' , and hence are disjoint in λ' .

Assume inductively that for $j \leq i - 1$, we have defined $\lambda'_x(s_j)$ such that $\lambda'_x(s_j) \in [b, b + (j - 1) \cdot \delta]$ for some $0 \leq b \leq j - 1$. We now place s_i . Set $\lambda'_x(s_i) = \infty$ initially. Consider the y -projections of s_i and s_{i-1} in λ . If they are not disjoint, then set $\lambda'_x(s_i) = \lambda'_x(s_{i-1}) + 1$. If not, s_i can be translated to the left till either $\lambda'_x(s_i) = \lambda'_x(s_{i-1}) + \delta$ or it touches some square in $s_{i'}, i' < i$, whichever happens first. In the former case, $\lambda'_x(s_i) \in [b, b + (i - 1) \cdot \delta]$ and $b \leq i - 1$. In the latter case, $s_{i'}$ satisfies the induction hypothesis and hence $\lambda'_x(s_i) (= \lambda'_x(s_{i'}) + 1)$ is in the interval $[b' + 1, b' + 1 + (i' - 1) \cdot \delta]$, where $b' \leq i' - 1 \leq i - 1$. Hence, in all cases, $\lambda'_x(s_i)$ satisfies the induction claim, and hence we claim it is true for all $i \leq n$.

It remains to bound the total width of the layout λ' . Since $\lambda'_x(s_1) = 1$ and $\lambda'_x(s_i) \leq b + i \cdot \delta$ for some integer $b \leq i - 1$, it follows that the width of the layout is in $[b, b + n \cdot \delta]$ for some integer b . Setting $\delta = \frac{\varepsilon}{n}$ concludes the proof. \square

Lemma 4. *For any $\varepsilon < 1/2$, if there is a disjoint layout λ' that has the same orthogonality as λ^{in} and whose height and width is bounded by $h + \varepsilon$ and $w + \varepsilon$ respectively, where h, w are positive integers, then there is a set of lines L that hit all segments in Φ , containing at most c vertical lines and r horizontal lines. Here $r = h - 1$, $c = w - 1$.*

Proof. Assume without loss of generality that λ' is contained in a rectangle $[0, w + \varepsilon] \times [0, h + \varepsilon]$ in the plane. We consider the following set of lines:

$$x = i + \frac{1}{2} + \delta, \text{ for } i = 1, 2, \dots, w - 1 \quad \text{and} \quad y = j + \frac{1}{2} + \delta, \text{ for } j = 1, 2, \dots, h - 1.$$

We pick a $\delta < \frac{1}{2}$ so that none of these lines contain a rectangle center. We now show that the resultant set of lines L constitute a hitting set for Φ .

Let $\phi \in \Phi$ be an arbitrary line segment, whose endpoints p, p' are the centers of unit squares s, s' in the disjoint layout λ' . Since s, s' do not intersect, then either $|\lambda'_x(s) - \lambda'_x(s')| \geq 1$ or $|\lambda'_y(s) - \lambda'_y(s')| \geq 1$. We assume without loss of generality its the former. We know that $0 \leq \lambda'_x(s), \lambda'_x(s') \leq w + \varepsilon$ and none of the vertical lines in L pass through p, p' . Since successive vertical lines are unit distance apart, there must be at least one vertical line $x = \frac{i'}{2} + \delta$ that lies between p, p' . We can argue similarly if $|\lambda'_y(s) - \lambda'_y(s')| \geq 1$ using the set of horizontal lines in L . Thus, Φ is hit by at least one line in L , making the latter a hitting set consisting of at most $c = w - 1$ vertical lines and $r = h - 1$ horizontal lines. \square

A.2 Proof of Theorem 6

Theorem 6. *There is no polynomial time $(1 + \xi)$ -factor approximation algorithm for UHS with $\xi \leq \frac{1}{55}\gamma$, unless $\mathbb{P} = \mathbb{NP}$, γ being the constant in Theorem 5.*

Proof. Suppose there is a polynomial time $(1 + \xi)$ -factor approximation algorithm for UHS for some $\xi \leq \frac{1}{55}\gamma$. Now consider the reduction π . For the instance of UHS obtained from an instance of 5-OCC-MAX-3SAT in which all the clauses are satisfied, there is a hitting set of size at most k ; so the approximation algorithm finds a hitting set of size at most $(1 + \xi)k$. For any instance obtained from an instance of 5-OCC-MAX-3SAT in which less than $1 - \gamma$ fraction of the clauses are satisfied, needs more than $(1 + \frac{1}{55}\gamma)k \geq (1 + \xi)k$ lines. Thus using this algorithm we can distinguish between an instance of 5-OCC-MAX-3SAT consisting of the clauses all of which can be satisfied, and one in which less than $1 - \gamma$ fraction of the clauses can be satisfied. Hence from Theorem 5 it follows that $\mathbb{P} = \mathbb{NP}$, which completes the proof of this theorem. \square

A.3 Reduction of 5-OCC-MAX-3SAT into Hitting Set Problem

A set L of horizontal and vertical lines is said to *separate* a set P of points, if for each point $p \in P$, there is a 2-dimensional cell (possibly unbounded) in the arrangement of the lines in L that contains p and no other points of P .

Recall that for any set of points P , $\Phi(P)$ is the set of segments induced by P . Suppose a set L of horizontal and vertical lines hits all the segments of $\Phi(P)$. By definition a line hits a segment, if it passes through the interior of the segment, but does not intersect either endpoint of the segment. Thus if we perturb each line in L that passes through a segment endpoint, then no line in L intersect any segment endpoint and L still remains a hitting set for $\Phi(P)$. Henceforth, by a hitting set of lines L for $\Phi(P)$ we mean the lines in L hit the segments in $\Phi(P)$, but no line in L intersects any point of P . Now our claim is that if L is a hitting set for $\Phi(P)$, L separates P . If not, then there exists a cell containing two points, and the corresponding segment is not being hit by any line, which cannot be true. Conversely, if L separates the points of P , then all the segments of $\Phi(P)$ are being hit. The notion of separation will help us to simplify our arguments for bounding the minimum number of lines required to hit all the segments induced by a point set. Thus from now onwards, we use the notion of hitting and separation interchangeably.

For any set of segments S and a hitting set of lines L for $S' \supseteq S$, consider the subset of L consisting of every line that hits at least one segment of S ; denote its cardinality by $N_L(S)$. For any set of segments S , denote the minimum number of lines needed to hit all segments of S by $N(S)$. Given a set of lines L , a point set P_1 is said to be separated from another point set P_2 , if each segment (p, p') is hit by some line in L , where $p \in P_1, p' \in P_2$. For any two point sets P_1 and P_2 , denote the minimum number of lines needed to separate P_1 from P_2 by $N(P_1, P_2)$.

For any set P of points, define x -span of P to be the interval $[x_{\min}, x_{\max}]$ on real line, where x_{\max} and x_{\min} are the maximum and minimum among the

x -coordinates of the points of P . Similarly, define y -span of P corresponding to the y -coordinates of its points. Now we have the following lemma.

Lemma 13. *Suppose P_1, \dots, P_l are point sets with pairwise disjoint x -spans and pairwise disjoint y -spans. Then for any hitting set of lines L for a set of segments $S \supseteq \bigcup_{i=1}^l \Phi(P_i)$,*

$$N_L\left(\bigcup_{i=1}^l \Phi(P_i)\right) = \sum_{i=1}^l N_L(\Phi(P_i))$$

Proof. Let L_i be the subset of L consisting of every line that hits at least one segment in $\Phi(P_i)$ for $i = 1, \dots, l$. Note that it is sufficient to show that for any $i, j \in \{1, \dots, l\}$ such that $i \neq j$, $L_i \cap L_j = \emptyset$. Consider any horizontal line $y = a$ of L_i . a must lie in y -span of P_i . As y -spans of P_i and P_j are disjoint, this line cannot hit any segment of $\Phi(P_j)$, and thus cannot belong to L_j . Similarly, any horizontal line of L_j cannot hit any segment of $\Phi(P_i)$, and thus cannot belong to L_i . Now consider any vertical line $x = b$ of L_i . b must lie in x -span of P_i . As x -spans of P_i and P_j are disjoint this line cannot hit any segment of $\Phi(P_j)$, and thus cannot belong to L_j . In a similar way, any vertical line of L_j cannot hit any segment of $\Phi(P_i)$, and thus cannot belong to L_i . Hence $L_i \cap L_j = \emptyset$. \square

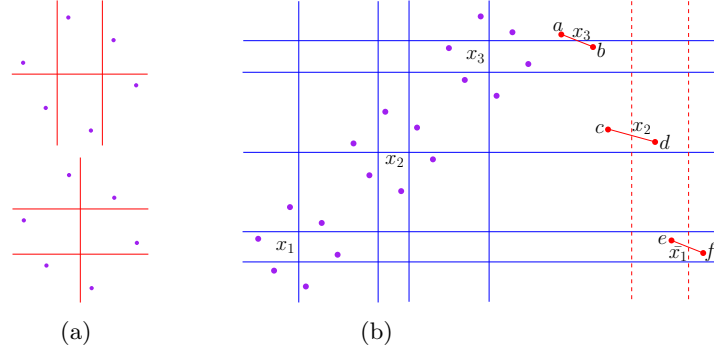


Fig. 1: (a) Optimal configurations of hitting set for variable (x_i) points. The top one corresponds to $x_i = 0$ and the bottom one corresponds to $x_i = 1$. (b) The points corresponding to the clause $x_3 \vee x_2 \vee \bar{x}_1$. The hitting set for variable gadgets corresponds to the assignment $x_1 = 1, x_2 = 0, x_3 = 1$

Now we discuss the construction of an instance of UHS from a given instance of 5-OCC-MAX-3SAT. We construct a gadget that consists of three classes of points - variable points, clause points and auxilliary points.

Variable Points For each variable we take a translate of the six points in $I = \{(1, 4), (2, 2), (3, 6), (4, 1), (5, 5), (6, 3)\}$. We note that three lines are necessary and sufficient to separate these six points. To be precise there are two

optimal choices one involving a horizontal line $y = a$, where $4 < a < 5$ and the other involving a horizontal line $y = b$, where $3 < b < 4$. See [Figure 1a](#) for an illustration. A set of points U' is called (p, q) -translate of a point set U , if U' is obtained by translating the x and y -coordinates of the points in U by p and q respectively. For each variable x_i we take a set V_i which is $((6(i-1), 6(i-1))$ -translate of I , where $1 \leq i \leq n$. Let $V = \cup_{i=1}^n V_i$.

We note that the x -spans (resp. y -spans) of the sets V_i for $1 \leq i \leq n$ are pairwise disjoint. With respect to V_i for $1 \leq i \leq n$, let us call a horizontal line a Type 0 (resp. Type 1) line, if it has the form $y = b$ for $6(i-1)+3 < b < 6(i-1)+4$ (resp. $y = a$ for $6(i-1)+4 < a < 6(i-1)+5$). Now we have the following observation.

Observation 4. *Three lines are necessary and sufficient to hit all the segments induced by V_i , i.e. $N(\Phi(V_i)) = 3$, where $1 \leq i \leq n$. In addition, there are exactly two types of optimal hitting sets for $\Phi(V_i)$ - one that uses a Type 1 line, but no Type 0 line, and the other that uses a Type 0 line, but no Type 1 line.*

We will associate the first choice of hitting set for $\Phi(V_i)$ with the assignment $x_i = 1$ and the second choice with $x_i = 0$. Now we have the following lemma.

Lemma 14. *$4n - 1$ lines are sufficient to hit all segments of $\Phi(V)$, where the segments of $\Phi(V_i)$ are being hit using 3 lines for $1 \leq i \leq n$.*

Proof. Note that separating V involves separating V_i from V_{i+1} for each $1 \leq i \leq n - 1$, and separating points of V_i for each $1 \leq i \leq n$. By [Observation 4](#) three lines are sufficient to separate the points of V_i . Also $N(V_i, V_{i+1}) = 1$, as any vertical or horizontal line between V_i and V_{i+1} separates one from the other. Thus in total $3n + n - 1 = 4n - 1$ lines are sufficient. \square

Given any hitting set for $\Phi(V)$ with at most $4n - 1$ lines we want to construct an assignment of the boolean formula ϕ . To be precise we want to ensure that always 3 lines are used to separate V_i , if at most $4n - 1$ lines are used to separate V . This ensures that one of the two configurations of 3 lines are used to separate V_i . Then we can assign binary values to the variables depending on the configuration used. [Figure 2](#) shows an example where $4n - 1$ lines are used to separate V , but 4 lines are used to separate one of the V_i . To ensure exactly 3 lines are used to separate each V_i , we can make sure that $n - 1$ lines between V_i and V_{i+1} are always picked up, for $1 \leq i \leq n - 1$. Then as 3 lines are needed to separate each V_i , exactly 3 lines are used for each V_i , if $4n - 1$ lines are used in total. Later we show that a subset of the auxilliary points will enforce this constraint.

Clause Points For each clause C_j we take a set T_j of ten points that is a union of two sets T_j^1 and T_j^2 . T_j^1 consists of a translate of the four points in $J = \{(2, -1), (5, -2), (6, -3), (9, -4)\}$. To be precise T_j^1 is $((6n + 1 + 10(j - 1), 4(j - m))$ -translate of J . In [Figure 3](#) p, q, r and s are the points of T_j^1 corresponding to the clause $x_3 \vee x_2 \vee \bar{x}_1$. The points in T_j^2 depend on the literals of C_j . Let

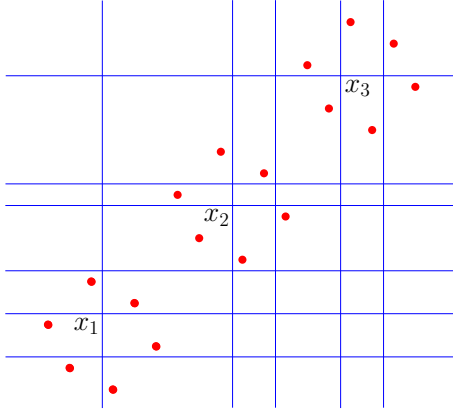


Fig. 2: 11 lines are used to separate V , but 4 lines are used to separate V_2

u_1, u_2 and u_3 are the literals of C_j . For each literal we take two points. The x -coordinates of the points corresponding to u_1, u_2 and u_3 are $\{6n+1+10(j-1)+1, 6n+1+10(j-1)+3\}$, $\{6n+1+10(j-1)+4, 6n+1+10(j-1)+7\}$ and $\{6n+1+10(j-1)+8, 6n+1+10(j-1)+10\}$ respectively. The y -coordinates of these points depend on the form of the literals. Let u_t is corresponding to the variable x_i for $t \in \{1, 2, 3\}$. If $u_t = x_i$, the y -coordinates of the two points corresponding to u_t are $6(i-1)+5-\epsilon j$ and $6(i-1)+4+\epsilon j$ respectively, where ϵ is a very small positive number. Otherwise, $u_t = \bar{x}_i$ and the y -coordinates are $6(i-1)+4-\epsilon j$ and $6(i-1)+3+\epsilon j$ respectively. Let $T^1 = \cup_{j=1}^m T_j^1$, $T^2 = \cup_{j=1}^m T_j^2$ and $T = T^1 \cup T^2$.

In Figure 3 the points corresponding to the literals x_3, x_2 and \bar{x}_1 are $\{a, b\}$, $\{c, d\}$ and $\{e, f\}$ respectively. Consider the points of T_j in decreasing order of y -coordinates. We divide the points in this order into groups of two. Thus the first three groups are corresponding to T_j^2 and the remaining two groups are corresponding to T_j^1 . We denote the first three groups by $T_{j,s}^2$ for $1 \leq s \leq 3$, and the remaining two groups by $T_{j,t}^1$ for $1 \leq t \leq 2$. We note that the y -spans of all these five groups are pairwise disjoint. Also x -spans of the three groups of T_j^2 are pairwise disjoint. Similarly, x -spans of the two groups of T_j^1 are pairwise disjoint. But the x -spans of $T_{j,1}^1$ and $T_{j,1}^2$, and $T_{j,1}^1$ and $T_{j,2}^2$ have nonempty intersection. Thus it is possible to separate the points in $T_{j,1}^1$ as well as $T_{j,1}^2$ (resp. $T_{j,1}^1$ as well as $T_{j,2}^2$) using a single vertical line. Similarly, it is possible to separate the points in $T_{j,2}^1$ as well as $T_{j,2}^2$ (resp. $T_{j,2}^1$ as well as $T_{j,3}^2$) using a single vertical line. See Figure 3 for an illustration. We note that x -spans (resp. y -spans) of the sets $T_{j,t}^1$ for $1 \leq t \leq 2$ and $1 \leq j \leq m$ are pairwise disjoint.

While constructing the points for C_1 we make sure that a literal corresponding to V_1 is never chosen as the first literal of C_1 . Note that this does not violate any generality, as each clause contains three distinct literals. Later this will help us to argue about the number of lines necessary to separate all the points.

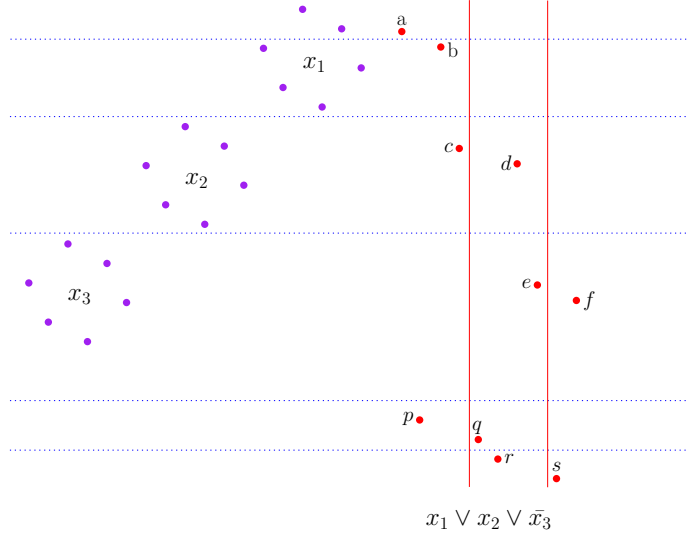


Fig. 3: The points corresponding to the clause $x_3 \vee x_2 \vee \bar{x}_1$ and a scheme to separate them

Now we discuss a scheme to separate the points of T_j using horizontal and vertical lines. The scheme considers the points in groups of two as defined before. At first we will see how to separate the groups from each other and then we will separate the points in the individual groups. Note that four horizontal lines are sufficient to separate these five groups of points from each other (see Figure 3). Also suppose the points in at least one group corresponding to the literals are separated by one additional horizontal line. Then two lines are necessary to separate the remaining points of T_j and two vertical lines are sufficient for that purpose. Hence we have the following observation.

Observation 5. *Suppose the five groups in T_j as defined above are separated from each other by four horizontal lines and points in at least one of the groups of T_j^2 are separated by a horizontal line, then two more lines are necessary to separate all the points of T_j and two vertical lines are sufficient for that purpose.*

Suppose that as a precondition we ensure that the five groups of T_j are separated from each other for each $1 \leq j \leq m$. Furthermore, suppose that we have fixed an optimal hitting set type for each V_i for $1 \leq i \leq n$. If the corresponding assignment satisfies the clause C_j , we can argue that the hitting sets for the $\Phi(V_i)$ have a horizontal line that separates the points in at least one of the three groups of T_j^2 . Then using Observation 5 two additional vertical lines suffice to separate the points in T_j . If the corresponding assignment does not satisfy C_j , then the points in none of the three groups of T_j^2 are separated by the optimal hitting sets for the $\Phi(V_i)$. Then at least three more lines are needed to separate the points in T_j .

Now we discuss how the preconditions could be satisfied. The three groups of T_j^2 can be separated from each other by the horizontal lines which are used to separate V_i and V_{i+1} for $1 \leq i \leq n$. An additional horizontal line would be sufficient to separate T_j^1 from T_j^2 for all $1 \leq j \leq m$. Now if a vertical line separates the two groups of T_j^1 from each other, then it also separates the points in the middle group of T_j^2 . Thus we cannot argue that two more lines are needed to separate the points of T_j . Hence we have to ensure that the two groups of T_j^1 are separated using a horizontal line. We will enforce this constraint by a subset of auxilliary points.

Auxilliary points The set of auxilliary points A consists of five point sets $\{A^1, \dots, A^5\}$. A^1 and A^2 ensure that $T_{j,1}^1$ and $T_{j,2}^1$ are separated by a horizontal line for $1 \leq j \leq m$. A^3 ensures that the points in each V_i are separated using exactly 3 lines, when $4n - 1$ lines are used to separate the points in V . A^4 ensures that T_j is separated from T_{j+1} by a vertical line for each $1 \leq j \leq m$. A^5 consists of four points $(6n + .5, .5), (6n + .5, -.5), (6n + 1.5, .5)$ and $(6n + 1.5, -.5)$. Note that 2 lines are necessary and sufficient to separate these four points. Moreover, a horizontal and a vertical line that separate these points also separate T^1 from T^2 , and V from T respectively (see Figure 4).

A^1 is composed of m point sets each of which consists of a translate of the four points in $F = \{(0, -\epsilon), (-1, -1 + \epsilon), (-2, -2\epsilon), (-3, -1 + 2\epsilon)\}$, where $\epsilon > 0$ is a very small number. $A^1 = \cup_{j=1}^m A_j^1$, where each A_j^1 is $(-4m + 3 + 4(j - 1), -4m + 2 + 4(j - 1))$ -translate of F (see Figure 4). Note that two lines are necessary, and one horizontal and one vertical line are sufficient to separate the points in each A_j^1 . By construction, the horizontal line corresponding to A_j^1 also separates the two groups of T_j^1 from each other (see Figure 4).

A^2 is composed of $m - 1$ point sets each of which consists of a translate of the four points in $H = \{(-\epsilon, 0), (-1 + \epsilon, -1), (-2\epsilon, -2), (-1 + 2\epsilon, -3)\}$. $A^2 = \cup_{l=1}^{m-1} A_l^2$, where each A_l^2 is $(-4m + 4l, -8m + 7 + 4(l - 1))$ -translate of H (see Figure 4). Note that one horizontal and one vertical line are sufficient to separate the points of each A_l^2 . Moreover, the vertical line corresponding to A_l^2 also separates the two groups A_l^1 and A_{l+1}^1 for $1 \leq l \leq m - 1$. And, any vertical line that hits a segment in $\Phi(A_l^1)$ separates A_{l-1}^2 from A_l^2 for $2 \leq l \leq m - 1$.

A^3 consists of $n - 1$ point sets each of which is a translate of the points in F . $A^3 = \cup_{l=1}^{n-1} A_l^3$, where A_l^3 is $(-4m - 1 - 4(l - 1), 7 + 6(l - 1))$ -translate of F . Thus the points in A^3 lie towards north-west of the points in A^1 (see Figure 4). Note that one horizontal and one vertical line are sufficient to separate the points in each A_l^3 . Also the y -span of A_l^3 is between the y -spans of V_l and V_{l+1} for each $1 \leq l \leq n - 1$. Thus the horizontal line corresponding to A_l^3 also separates V_l from V_{l+1} for $1 \leq l \leq n - 1$.

A^4 consists of $m - 1$ point sets each of which is a translate of the points in H . $A^4 = \cup_{j=1}^{m-1} A_j^4$, where A_j^4 is $(-8m + 3 - 4(j - 1), 6n + 2 + 10j)$ -translate of H . Note that two lines are necessary, and one horizontal and one vertical line are sufficient to separate the points in each A_j^4 . In addition, the x -span of A_j^4 is between the x -spans of T_j and T_{j+1} for each $1 \leq j \leq m - 1$. Thus the vertical

line corresponding to A_j^4 also separates T_j from T_{j+1} for $1 \leq j \leq m-1$ (see Figure 4).

Now consider the collection of sets A_l^1 for $1 \leq l \leq m$, A_s^2 for $1 \leq s \leq m-1$, A_t^3 for $1 \leq t \leq n-1$, A_j^4 for $1 \leq j \leq m-1$, $T_{j,1}^1$ for $1 \leq j \leq m$ and $T_{j,2}^1$ for $1 \leq j \leq m$. By construction, x -spans (resp. y -spans) of any two sets in this collection are disjoint. Also y -spans (resp. x -spans) of the sets V_i for $1 \leq i \leq n$ and A_t^3 for $1 \leq t \leq n-1$ are pairwise disjoint. Similarly, y -spans (resp. x -spans) of the sets T_j^1 for $1 \leq j \leq m$ and A_j^4 for $1 \leq j \leq m-1$ are pairwise disjoint. As x -spans (resp. y -spans) of V and $A^1 \cup A^2$, V and $A^4 \cup A^5$, and V and $\cup_{j=1}^m T_j^1$ are disjoint, we have the following observation.

Observation 6. *Consider the collection of sets A_t^1 for $1 \leq t \leq m$, A_l^2 for $1 \leq l \leq m-1$, A_s^3 for $1 \leq s \leq n-1$, A_j^4 for $1 \leq j \leq m-1$, V_i for $1 \leq i \leq n-1$, $T_{j,1}^1$ for $1 \leq j \leq m$, $T_{j,2}^1$ for $1 \leq j \leq m$ and A^5 . x -spans (resp. y -spans) of any two sets in this collection are disjoint.*

Now we have the following lemma regarding the properties of the auxilliary points.

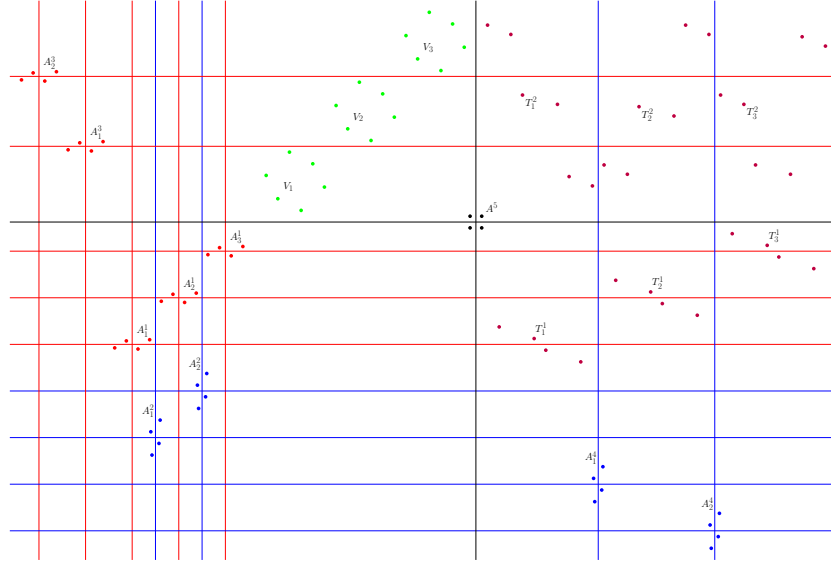


Fig. 4: Construction of points corresponding to the clauses $\{x_3 \vee x_2 \vee x_1, x_1 \vee x_2 \vee x_3, x_2 \vee x_1 \vee x_3\}$ ($n = 3, m = 3$).

Lemma 15. *The points of A have the following properties:*

- i) $4m - 2$ lines are necessary to hit the segments in $\Phi(A^1 \cup A^2)$. Moreover, there is a hitting set for $\Phi(A^1 \cup A^2)$ consisting of $4m - 2$ lines that also separates the two groups of T_j^1 from each other for $1 \leq j \leq m$.

- ii) $2(n-1)$ lines are necessary and sufficient to hit the segments in $\bigcup_{t=1}^{n-1} \Phi(A_t^3)$.
iii) $2(m-1)$ lines are necessary and sufficient to hit the segments in $\bigcup_{j=1}^{m-1} \Phi(A_j^4)$.

Proof. i) Consider any hitting set of lines L for the segments in $\Phi(A^1 \cup A^2)$. At first we show that $N_L(\Phi(A^1 \cup A^2)) \geq 4m - 2$. Note that $N(\Phi(A_j^1)) = N(\Phi(A_l^2)) = 2$ for $1 \leq j \leq m, 1 \leq l \leq m-1$. Also x -spans and y -spans of the sets A_j^1 for $1 \leq j \leq m$ and A_l^2 for $1 \leq l \leq m-1$ are pairwise disjoint. Hence by [Lemma 13](#), $N_L(\Phi(A^1 \cup A^2)) \geq N_L(\bigcup_{j=1}^m \Phi(A_j^1) \cup \bigcup_{l=1}^{m-1} \Phi(A_l^2)) = \sum_{j=1}^m N_L(\Phi(A_j^1)) + \sum_{l=1}^{m-1} N_L(\Phi(A_l^2)) \geq \sum_{j=1}^m N(\Phi(A_j^1)) + \sum_{l=1}^{m-1} N(\Phi(A_l^2)) = 2(m+m-1) = 4m-2$.

Now we construct a hitting set of lines for $\Phi(A^1 \cup A^2)$ having at most $4m-2$ lines. We take one horizontal and one vertical line to separate the points of each of the sets A_j^1 and A_l^2 for $1 \leq j \leq m, 1 \leq l \leq m-1$. We need $4m-2$ lines for this. Now the vertical line corresponding to A_l^2 also separates A_l^1 from A_{l+1}^1 for $1 \leq l \leq m-1$. And, the vertical line corresponding to A_j^1 also separates the groups A_{j-1}^2 and A_j^2 for $2 \leq j \leq m-1$. Thus all the points in $A^1 \cup A^2$ are separated. By construction, the horizontal line corresponding to A_j^1 also separates the two groups of T_j^1 from each other, and the result follows.

ii) First we prove that $2(n-1)$ lines are necessary. Consider any hitting set of lines L for the segments in $\bigcup_{t=1}^{n-1} \Phi(A_t^3)$. We note that $N(\Phi(A_t^3)) = 2$ for $1 \leq t \leq n-1$. Also x -spans and y -spans of the sets A_t^3 for $1 \leq t \leq n-1$ are pairwise disjoint. Thus using [Lemma 13](#),

$$N_L(\bigcup_{t=1}^{n-1} \Phi(A_t^3)) = \sum_{t=1}^{n-1} N_L(\Phi(A_t^3)) \geq \sum_{t=1}^{n-1} N(\Phi(A_t^3)) = 2(n-1).$$

To prove the sufficient condition we construct a hitting set of lines for the segments in $\bigcup_{t=1}^{n-1} \Phi(A_t^3)$. For each $1 \leq t \leq n-1$ we take one horizontal and one vertical line that separates the points in A_t^3 . Hence the $2(n-1)$ lines in total hit all the segments in $\bigcup_{t=1}^{n-1} \Phi(A_t^3)$.

iii) Consider any hitting set of lines L for the segments in $\bigcup_{j=1}^{m-1} \Phi(A_j^4)$. We note that $N(\Phi(A_j^4)) = 2$ for $1 \leq j \leq m-1$. Also x -spans and y -spans of the sets A_j^4 for $1 \leq j \leq m-1$ are pairwise disjoint. Thus using [Lemma 13](#),

$$N_L(\bigcup_{j=1}^{m-1} \Phi(A_j^4)) = \sum_{j=1}^{m-1} N_L(\Phi(A_j^4)) \geq \sum_{j=1}^{m-1} N(\Phi(A_j^4)) = 2(m-1).$$

Now to prove the sufficient condition we construct a hitting set for the segments in $\bigcup_{t=1}^{m-1} \Phi(A_t^4)$ consisting of $2m-2$ lines. For each $1 \leq j \leq m-1$ we take one horizontal and one vertical line that separates the points in A_j^4 . Hence the $2(m-1)$ lines in total hit all the segments in $\bigcup_{j=1}^{m-1} \Phi(A_j^4)$. \square

Now we formally discuss the reduction from 5-OCC-MAX-3SAT to *Hitting Set Problem*. Given a 5-OCC-MAX-3SAT formula ϕ we construct the set of points V and T corresponding to variables and clauses respectively in the way mentioned before. We also construct A , the set of auxilliary points, in the same way mentioned before. Let $P = V \cup T \cup A$. Let $k = 5n + 8m - 4$. Now we show that the reduction has the two properties as claimed before. The following lemma ensures the first property.

Lemma 16. *Suppose ϕ is satisfiable, then $N(\Phi(P)) \leq k$.*

Proof. Note that it is sufficient to show the existence of a set L of at most k horizontal and vertical lines which separate the points of P . The set L is constructed in the following way. L is composed of three disjoint sets of lines L^v, L^c and L^a . L^v, L^c and L^a are the sets of lines which separate the points of V, T and A respectively. These lines also separate the groups V, T and A from each other. Now we will consider the construction of these sets.

We add the horizontal line $y = 0$ and the vertical line $x = 6n + 1$ to L^a . These two lines divide the plane into four cells. Each cell contains one point of A^5 and some additional points. Thus the points in A^5 are separated by these two lines. In addition to the points of A^5 the four cells contain the points of four disjoint sets $A^1 \cup A^2, V \cup A^3, T^1 \cup A^4$, and T^2 respectively. Hence the four point sets in the four cells are already separated from each other. We show how to separate the points in each of these four sets.

By Lemma 15(i) it follows that there is a hitting set for $\Phi(A^1 \cup A^2)$ consisting of $4m - 2$ lines that also separates the two groups of T_j^1 from each other for $1 \leq j \leq m$. We add such $4m - 2$ lines to L^a . Now $|L^a| = 4m$.

We note that the vertical line corresponding to A_1^1 separates A^3 from V . By Lemma 15(ii) $2(n - 1)$ lines are sufficient to hit the segments in $\Phi(A^3)$. We add such a set of $2n - 2$ lines to L^a . Now $|L^a| = 4m + 2n - 2$. Note that the horizontal line corresponding to A_i^3 separates V_i from V_{i+1} for $1 \leq i \leq n - 1$. By Observation 4 three lines are sufficient to separate the points of V_i and only two configurations of lines are possible for that purpose. For each variable x_i , if x_i is 1, we select the configuration with the Type 1 horizontal line $y = 6(i - 1) + 4.5$ for V_i . If x_i is 0, we select the configuration with the Type 0 horizontal line $y = 6(i - 1) + 3.5$ for V_i . We add the 3 lines corresponding to the chosen configuration to L^v for each V_i . Thus $|L^v| = 3n$. Hence the points in V are now separated. Also at least one horizontal line has been chosen per V_i and thus A_t^3 is separated from A_{t+1}^3 for each $1 \leq t \leq n - 2$. Thus all the points in $V \cup A^3$ are now being separated. Note that a vertical line corresponding to V_n and a horizontal line corresponding to V_1 separate $V \cup A^3$ from A^4 . Also the same vertical line separates $A^1 \cup A^2$ from A^4 .

By Lemma 15(iii) $2(m - 1)$ lines are sufficient to hit the segments in $\Phi(A^4)$. We add such a set of $2m - 2$ lines to L^a . Now $|L^a| = 4m + 2n - 2 + 2m - 2 = 6m + 2n - 4$. Note that the vertical line corresponding to A_j^4 also separates T_j from T_{j+1} for all $1 \leq j \leq m - 1$. Also the horizontal line corresponding to A_1^2 separates A^4 from T_1 .

Now consider the points in T . The horizontal lines that separate V_i from V_{i+1} for all $1 \leq i \leq n-1$, also separate the three groups $T_{j,1}^2, T_{j,2}^2$ and $T_{j,1}^2$ from each other, for $1 \leq j \leq m$. Also the horizontal lines in L^a corresponding to A^1 separate $T_{j,1}^1$ from $T_{j,2}^1$ for $1 \leq j \leq m$. Now we claim that for each $1 \leq j \leq m$ the points in at least one of the groups of T_j^2 are being separated by a horizontal line in L^v . Consider any clause $C_j = u_1 \vee u_2 \vee u_3$. Now as C_j is satisfied there must be a satisfied literal of C_j . Without loss of generality let u_1 be a satisfied literal of C_j and it is corresponding to the variable x_i . If u_1 is x_i , the points in T_j^2 corresponding to u_1 are $(6n+1+10(j-1)+1, 6(i-1)+5-\epsilon j)$ and $(6n+1+10(j-1)+3, 6(i-1)+4+\epsilon j)$. As x_i is 1, previously we chose the configuration with the Type 1 line for V_i . Thus these two points are already separated. Otherwise, if u_1 is \bar{x}_i , the points in T_j^2 corresponding to u_1 are $(6n+1+10(j-1)+1, 6(i-1)+4-\epsilon j)$ and $(6n+1+10(j-1)+3, 6(i-1)+3+\epsilon j)$. Now as x_i is 0, previously we chose the configuration with the Type 0 line for V_i . Thus in this case also the two points are already separated.

Consider the set of points T_j for any $1 \leq j \leq m$. All the five groups of T_j are separated from each other, and the points in at least one group of T_j^2 are separated. Thus by **Observation 5** only two vertical lines are sufficient to separate the points in T_j . We construct such a set L^c containing $2m$ vertical lines corresponding to T_j for all $1 \leq j \leq m$. As the points in T_j are separated for all $1 \leq j \leq m$, and T_j is separated from T_{j+1} for all $1 \leq j \leq m$, all the points in T are separated.

Also the vertical line corresponding to $T_{1,1}^1$ and the horizontal line corresponding to A_m^1 separate A^4 from T^1 . Now the first literal of C_1 cannot be corresponding to V_1 . Thus the vertical line corresponding to $T_{1,1}^1$ and a horizontal line corresponding to V_1 separate A^4 from T^2 . Lastly, the vertical line corresponding to $T_{j,1}^1$ separate A_{j-1}^1 from A_j^1 for $2 \leq j \leq m-1$.

We set $L = L^a \cup L^v \cup L^c$ and thus the lines of L separate all the points of P by our construction. Now $|L| = (6m+2n-4) + 3n+2m = 5n+8m-4 = k$ which completes the proof of the lemma. \square

Now we show that the reduction ensures the second property as well, i.e if less than $1-\delta$ ($0 < \delta \leq 1$) fraction of the clauses of ϕ are satisfied, then more than $(1+\frac{1}{55}\delta)k$ lines are needed to hit the segments in $\Phi(P)$. The idea is to show that any hitting set of lines for $\Phi(P)$ that uses exactly 3 lines to separate the points in V_i , uses at least k lines. Thus if at most $k+p$ lines are used to hit all segments in $\Phi(P)$, then more than 3 lines can be used for at most p V_i 's. Now if 3 lines are used to separate the points in V_i , then we can find a binary assignment of the variable. Using the exact three lines configurations of at least $n-p$ remaining V_i 's we can create an assignment (the values of the other variables are chosen arbitrarily) which makes a good fraction of the clauses to be satisfied. Before moving on now we have some definitions.

Consider any hitting set of lines L for $\Phi(P)$. A variable x_i is said to be *good*, if the number of lines of L that hit at least one segment of $\Phi(V_i)$ is equal to 3. A variable is said to be *bad*, if it is not good. A clause C_j is said to be *bad*, if it has at least one of the following two properties, (i) any line of L that hits the

segment of $\Phi(T_{j,t}^2)$, is vertical, for each $1 \leq t \leq 3$, and (ii) it contains a literal corresponding to a bad variable. The bad clauses which have the first property are called bad clauses of first type, and the remaining bad clauses are called bad clauses of second type. A clause is said to be *good*, if it is not bad. Now we have the following lemma.

Lemma 17. *Consider any hitting set L for $\Phi(P)$ consisting of at most $k + p$ lines. The number of bad clauses corresponding to L is at most $5p$.*

Proof. Let a be the number of good variables. Then

$$\sum_{i=1}^n N_L(\Phi(V_i)) \geq 3a + 4(n - a) = 4n - a \quad (14)$$

As each variable appears in exactly 5 clauses the bad variables can appear in at most $5(n - a)$ clauses. Let b be the number of bad clauses of first type. Then the total number of bad clauses is at most $5(n - a) + b$. Now consider a bad clause C_l of first type. A vertical line that separate the points in $T_{l,t}^2$ might not separate the points of any of $T_{l,1}^1$ and $T_{l,2}^1$. Let C' be the subset of the bad clauses which has one such corresponding vertical line, and denote the cardinality of C' by b_1 . Then each of the three vertical lines, corresponding to any of the remaining $b - b_1$ bad clauses C_s separate the points in either $T_{s,1}^1$ or $T_{s,2}^1$. Now for any clause C_j x -spans (resp. y -spans) of the sets $T_{j,1}^1$ and $T_{j,2}^1$ are disjoint. Thus at least two lines are needed to separate the points in $T_{j,1}^1$ and $T_{j,2}^1$. This implies

$$\sum_{j=1}^m N_L(\Phi(T_{j,1}^1)) + \sum_{j=1}^m N_L(\Phi(T_{j,2}^1)) \geq 2m + (b - b_1) \quad (15)$$

By [Observation 6](#) the sets A_t^1 for $1 \leq t \leq m$, A_l^2 for $1 \leq l \leq m - 1$, A_s^3 for $1 \leq s \leq n - 1$, A_j^4 for $1 \leq j \leq m - 1$, V_i for $1 \leq i \leq n - 1$, $T_{j,1}^1$ for $1 \leq j \leq m$, $T_{j,2}^1$ for $1 \leq j \leq m$ and A^5 have pairwise disjoint x spans and pairwise disjoint y -spans. Applying [Lemma 13](#), [Equation \(14\)](#) and [Equation \(15\)](#) we count a lower bound of $N_L(\Phi(P))$ which is as follows.

$$\begin{aligned}
& N_L(\Phi(P)) \\
& \geq N_L\left(\bigcup_{t=1}^m \Phi(A_t^1) \cup \bigcup_{l=1}^{m-1} \Phi(A_l^2) \cup \bigcup_{s=1}^{n-1} \Phi(A_s^3) \cup \bigcup_{j=1}^{m-1} \Phi(A_j^4) \cup \bigcup_{i=1}^n \Phi(V_i) \cup \right. \\
& \quad \left. \bigcup_{j=1}^m \Phi(T_{j,1}^1) \cup \bigcup_{j=1}^m \Phi(T_{j,2}^1) \cup \Phi(A^5)\right) \\
& = \sum_{t=1}^m N_L(\Phi(A_t^1)) + \sum_{l=1}^{m-1} N_L(\Phi(A_l^2)) + \sum_{s=1}^{n-1} N_L(\Phi(A_s^3)) + \sum_{j=1}^{m-1} N_L(\Phi(A_j^4)) + \\
& \quad \sum_{i=1}^n N_L(\Phi(V_i)) + \sum_{j=1}^m N_L(\Phi(T_{j,1}^1)) + \sum_{j=1}^m N_L(\Phi(T_{j,2}^1)) + N_L(\Phi(A^5)) \\
& \geq \sum_{t=1}^m N(\Phi(A_t^1)) + \sum_{l=1}^{m-1} N(\Phi(A_l^2)) + \sum_{s=1}^{n-1} N(\Phi(A_s^3)) + \sum_{j=1}^{m-1} N(\Phi(A_j^4)) + \\
& \quad (4n - a) + (2m + b - b_1) + N(\Phi(A^5)) \\
& = \sum_{t=1}^m 2 + \sum_{l=1}^{m-1} 2 + \sum_{s=1}^{n-1} 2 + \sum_{j=1}^{m-1} 2 + (4n - a) + (2m + b - b_1) + 2 \\
& = 2m + 2m - 2 + 2n - 2 + 2m - 2 + 4n - a + 2m + b - b_1 + 2 \\
& = (8m + 5n - 4) + (n + b - a - b_1) = k + (n + b - a - b_1).
\end{aligned}$$

Now consider any clause C_j in C' . There is at least one vertical line l_j in L that separate the points in one of the groups of T_j^2 , but do not separate the points in any of $T_{j,1}^1$ and $T_{j,2}^1$. Let l_j separates the points in $T_{j,t}^2$ for some $1 \leq t \leq 3$. The x -coordinates of the points in $T_{j,t}^2$ are $6n + 1 + 10(j - 1) + 3(t - 1) + 1$ and $6n + 1 + 10(j - 1) + 3(t - 1) + 3$. Then the equation of l_j is $x = c$ for some $6n + 1 + 10(j - 1) + 3(t - 1) + 1 < c < 6n + 1 + 10(j - 1) + 3(t - 1) + 3$. We claim that the vertical line l_j is not considered in the previous counting of lower bound of $N_L(\Phi(P))$. If not, then one of the groups considered in the previous counting must contain two points (p_x^1, p_y^1) and (p_x^2, p_y^2) such that $p_x^1 < c$ and $p_x^2 > c$. The maximum among the x -coordinates of the points in $A \setminus A_4$ is $6n + 1.5 < c$, which is actually a point of A^5 . Thus none of the subsets of $A \setminus A_4$ contain such points. Also the x -spans of A_t^4 and T_j are disjoint for any $1 \leq t \leq m - 1$. Thus no such A_t^4 contains those points. Similarly, the maximum among the x -coordinates of the points in V is $6n < c$, and thus no V_i contains such points. Now consider any T_l^1 for $l < j$. Then the maximum among the x -coordinates of the points in T_l^1 is $6n + 1 + 10(l - 1) + 9 = 6n + 1 + 10l - 1 \leq 6n + 1 + 10(j - 1) - 1 \leq 6n + 1 + 10(j - 1) + 3(t - 1) + 1 < c$. Thus none of $T_{l,1}^1$ and $T_{l,2}^1$ contain such points. Now consider any T_l^1 for $l > j$. Then the minimum among the x -coordinates of the points in T_l^1 is $6n + 1 + 10(l - 1) + 2 \geq 6n + 1 + 10(j + 1 - 1) + 2 \geq 6n + 1 + 10(j - 1) + 10 + 2 \geq 6n + 1 + 10(j - 1) + 3(3 - 1) + 6 > 6n + 1 + 10(j - 1) + 3(t - 1) + 2 > c$. Thus

in this case also none of $T_{l,1}^1$ and $T_{l,2}^1$ contain the desired points. But this leads to a contradiction and hence l_j is not considered in the counting before.

Now $|C'| = b_1$ and thus in total there are b_1 additional vertical lines in L . So $N_L(\Phi(P)) \geq k + (n + b - a - b_1) + b_1 = k + n + b - a$. As the total number of lines in L is at most $k + p$,

$$k + n + b - a \leq k + p$$

or, $n - a + b \leq p$

Thus the total number of bad clauses is $5(n - a) + b \leq 5(n - a + b) \leq 5p$. \square

Lemma 18. *Suppose at most $k + \frac{\delta}{5}m$ ($0 < \delta \leq 1$) lines are sufficient to hit the segments in $\Phi(P)$, then there is an assignment which satisfy at least $1 - \delta$ fraction of the clauses.*

Proof. Consider any hitting set of lines L for $\Phi(P)$ which uses at most $k + \frac{\delta}{5}m$ lines. We construct an assignment of the clauses having the desired property. Let V^1 be the set of good variables corresponding to L . Then by **Observation 4** for each $x_i \in V^1$ the optimal configuration with either a Type 1 line or a Type 0 line has been chosen. For each $x_i \in V^1$ if the configuration with Type 1 line is chosen, set x_i to 1. Otherwise, set x_i to 0. For each $x_i \in V \setminus V^1$ assign an arbitrary binary value to x_i . We show that this assignment makes at least $(1 - \delta)m$ clauses to be satisfied.

Consider any good clause C_j . All of its literals are corresponding to good variables and the points of at most two groups of T_j^2 have been separated using only vertical lines. Let $T_{j,t}^2$ be the group whose points are separated by a horizontal line. Let u and x_i be the literal and variable corresponding to this group. If u is x_i , the y -coordinates of the points in $T_{j,t}^2$ are $6(i - 1) + 5 - \epsilon j$ and $6(i - 1) + 4 + \epsilon j$. Thus the horizontal line that separates these two points is of Type 1, which implies 1 is assigned to x_i and C_j is true. If u is \bar{x}_i , the y -coordinates of the points in $T_{j,t}^2$ are $6(i - 1) + 4 - \epsilon j$ and $6(i - 1) + 3 + \epsilon j$. Thus the horizontal line that separates the points in $T_{j,t}^2$ is of Type 0. This implies 0 is assigned to x_i and in this case also C_j is true. Thus any good clause is satisfied by the assignment. From **Lemma 17** it implies that the number of good clauses is at least $m - 5\frac{\delta}{5}m = (1 - \delta)m$, which completes the proof of this lemma. \square

Using the two equations $k = 5n + 8m - 4$ and $3m = 5n$ we get the following inequality.

$$m \geq \frac{1}{11}k \tag{16}$$

Combining the **Inequality (16)** with the contrapositive of **Lemma 18** it follows that, if less than $1 - \delta$ fraction of the clauses of ϕ are satisfied, then the number of lines needed is more than $k + \frac{1}{55}\delta k = (1 + \frac{1}{55}\delta)k$. Thus our reduction ensures both of the properties we claimed.

A.3.1 Proof of Theorem 8

Theorem 8. *Suppose there is a polynomial time algorithm that, given $\Phi(P)$ and non-negative integers r, c as input to CUHS,*

- (1) *outputs “yes”, if there is a set with at most c vertical and r horizontal lines that hits the segments in $\Phi(P)$; and*
- (2) *outputs “no”, if there is no hitting set for $\Phi(P)$ using at most $(1+\xi)c$ vertical and $(1+\xi)r$ horizontal lines, where ξ is the constant in Theorem 6.*

Then $\mathbb{P} = \mathbb{NP}$.

Proof. Assume that there is such an algorithm \mathbb{A} . Then we show that, given an instance (P, k) of UHS, it is possible to decide in polynomial time whether it can be hit by at most k lines, or any hitting set corresponding to it needs more than $(1+\xi)k$ lines. Thus from Theorem 6 it follows that $\mathbb{P} = \mathbb{NP}$.

Given the instance (P, k) of UHS, we call \mathbb{A} on P, r, c for each pair of positive integers r and c such that $r+c = k$. If \mathbb{A} returns yes for at least one pair, then we return “yes” for the instance of UHS. Otherwise, we return “no” for the instance of UHS.

If the instance (P, k) admits a hitting set of size at most k , fix such a hitting set L . Let L contains r' horizontal and c' vertical lines. Then for $r = r'$ and $c = k - c' \geq r'$, \mathbb{A} returns “yes”, and so we return “yes” for the instance of UHS. If the instance (P, k) does not admit a hitting set of size at most $(1+\xi)k$, then there is no hitting set for $\Phi(P)$ with at most $(1+\xi)r$ horizontal and $(1+\xi)c$ vertical lines such that $r+c = k$. Thus \mathbb{A} returns “no” on all (r, c) pairs on which it is invoked, and so we return “no” for the instance of UHS. \square

A.3.2 Proof of Theorem 9

Theorem 9. *Suppose there is a polynomial time algorithm that, given S, λ^{in} , w, h , and ε as input to CLADS,*

- (1) *outputs “yes”, if there is an output layout λ' with $H(\lambda') \leq h+\varepsilon$ and $W(\lambda') \leq w+\varepsilon$; and*
- (2) *outputs “no”, if there is no output layout λ' with $H(\lambda') \leq (1+\xi')(h+\varepsilon)$ and $W(\lambda') \leq (1+\xi')(w+\varepsilon)$, where $\xi' = \frac{\xi}{4}$ and ξ is the constant in Theorem 6.*

Then $\mathbb{P} = \mathbb{NP}$.

Proof. Assume that such an algorithm \mathbb{B} does exist. Then we show that, given an instance (P, r, c) of CUHS, it is possible to decide in polynomial time whether there is a set with at most c vertical and r horizontal lines that hits the segments in $\Phi(P)$, or there is no hitting set for $\Phi(P)$ using at most $(1+\xi)c$ vertical and $(1+\xi)r$ horizontal lines. By Theorem 8 this is true only if $\mathbb{P} = \mathbb{NP}$ and hence the result follows.

Given the instance (P, r, c) of CUHS, if $r = 0$ or $c = 0$, CUHS can be solved in polynomial time. Otherwise, we construct a set of squares S' by taking a unit

square centered at p for each point $p \in P$. We invoke \mathbb{B} on $S = S'$, $h = r + 1$, $w = c + 1$ and $\varepsilon = \xi'$. If \mathbb{B} returns “yes”, we return “yes”. Otherwise, we return “no”.

Suppose the instance (P, r, c) does admit a hitting set with at most c vertical and r horizontal lines. Then by [Lemma 1](#) there is an output layout λ' with $H(\lambda') \leq r + 1 + \xi'$ and width $W(\lambda') \leq c + 1 + \xi'$. Thus \mathbb{B} returns “yes”, and so we return “yes” for the instance of CUHS.

Suppose the instance (P, r, c) does not admit a hitting set with at most $(1 + \xi)c$ vertical and $(1 + \xi)r$ horizontal lines. Then by [Lemma 4](#), for any output layout λ' of the squares in S , either $H(\lambda') > (1 + \xi)r + 1$ or $W(\lambda') > (1 + \xi)c + 1$. Now as $c \geq 1$, $(1 + \xi)c + 1 = c + 1 + 4\xi'c = c + \xi'c + 1 + 2\xi'c + \xi'c > c + \xi'c + 1 + 2\xi' + \xi'^2 = (1 + \xi')(c + 1 + \xi')$. Similarly, as $r \geq 1$, $(1 + \xi)r + 1 > (1 + \xi')(r + 1 + \xi')$. Hence either $H(\lambda') > (1 + \xi')(r + 1 + \xi')$ or $W(\lambda') > (1 + \xi')(c + 1 + \xi')$. Thus at least one of the two conditions $W(\lambda') \leq (1 + \xi')(w + \varepsilon)$ and $H(\lambda') \leq (1 + \xi')(h + \varepsilon)$ is false, and \mathbb{B} returns “no”, and so we return “no”. \square